

*A1*

In addition, the procedures required by the exemplary editors are somehow tedious and laborious and can be inherently of high cost. Quite often, a business that has many documents to convert has to outsource the process due to the inefficiency and slowness associated with the conversion process. On the other end, the conversion process conducted by a service provider is difficult to be quantified as it is mainly involved in manual and repeated processes depending on the complexities of the documents. There is thus another need for a mechanism for quantifying the conversion of the unstructured documents to structured documents for various presentations in a cost-determinable way.

2. The following paragraph replaces the original paragraph that begins on page 15, line 14, and ends on page 15, line 24 of the specification:

*A2*

FIG. 2A illustrates an example of an unstructured document 200 that may be composed, edited or managed by an authoring tool. In an unstructured document, data is generally presented in sequence, which usually follows a reading order (e.g. from top to bottom and left to right). This sequence may be parsed into segments of data elements, where each data element 202 is assigned with decoration attributes or information such as positions, font color, font size, font type, style and various effects and etc. The decoration information is essentially for proper layout and presentation purpose when a file containing the data elements is opened by the authoring tool for display on a display screen.

3. The following paragraph replaces the original paragraph that begins on page 16, line 21, and ends on page 17, line 9 of the specification:

*A3*

Unlike the unstructured document, the structured document can easily access certain information via the document elements. Presentation of a structured document is usually defined in separate style sheets, e.g., written in cascading style sheet (CSS) or extensible style language for formatting objects (XSL-FO), which interprets layout for each document element. This feature allows a structured document to be presented in different layouts for different media through different style sheets. Generally, the decoration information or formatting attributes, such as font information in an unstructured document, unless defined in DTD as attributes of document elements, are abandoned after an unstructured document is converted into a corresponding structured

document. Further modification of formatting information will in general not affect the converted structured documents.

4. The following paragraph replaces the original paragraph that begins on page 17, line 10, and ends on page 17, line 21 of the specification:

*A4* FIG. 3A illustrates a functional diagram 300 according to one embodiment of the present invention. A conversion module 302 comprises an association module 304 and an integration module 306. Association module 304 receives an unstructured document, preferably in a metafile format. At the same time, association module 304 also receives a file, referred to as a definition file including DTD that are predefined. Generally, DTD is defined according to the nature or purposes of the unstructured document. For example, the unstructured document is in a category of receipts, e.g. document 200 in FIG. 2A, the DTD in a definition file as shown in FIG. 2B is designed in accordance to the "receipt-type" documents.

5. The following paragraph replaces the original paragraph that begins on page 17, line 22, and ends on page 18, line 18 of the specification:

*A5* To further understand association module 304, FIG. 3B shows an environment 320 implementing conversion module 302 according to one embodiment of the present invention. Environment 320 includes two displays 322 and 324 for a user to perform a conversion of an unstructured document to a file in markup language (referring to a markup language file). Display 322 is used to display the unstructured document. In one preferable embodiment, a metafile version of the unstructured document is loaded for display. A metafile, referring to either the unstructured document or a printed version thereof, typically contains many displayable objects. Each object is a cluster or a group of characters or words or a graphic representation. As shown in display 322, each word or an isolated numeral is a displayable object which is inherently carried over in the metafile. In other words, each object is defined by a number of attributes or decoration information including, but not limited to, type, size, color and position of the object such that it can be "printed" correctly. A number of objects can be grouped manually by a user in terms of their meanings or purposes. For example, group object 326 includes three character-type objects

"Green", "Chili" and "Salsa". Naturally the three character-type objects forms a title as a group object 326. The object grouping may be performed for the rest of the displayed metafile in display 322.

6. The following paragraph replaces the original paragraph that begins on page 18, line 19, and ends on page 18, line 24 of the specification:

A6 Display 324 is used to display a definition file prepared for the metafile in display 322. To facilitate operations of association module 304, the definition file is presented graphically as "DTD Pool" 328. For example, the graphical representation 328 of DTD 208 in FIG. 2B is used in display 324 to illustrate the hierarchical relationships among the document elements.

7. The following paragraph replaces the original paragraph that begins on page 22, line 4, and ends on page 22, line 10 of the specification:

A7 FIG. 3E shows a process flowchart 370 of using a product including an implementation of conversion module 302 according to one embodiment of the present invention. Sometimes, the product is leased by a user or a business. Other times, the product is used by a service provider providing services to businesses that need to convert unstructured documents to structured documents for different media presentation (e.g. presentation on a web site).

8. The following paragraph replaces the original paragraph that begins on page 26, line 7, and ends on page 26, line 16 of the specification:

A8 FIG. 6 shows an editing result 600 for the unstructured document 200 of FIG. 2A. Each parsed data element or combined objects 602, 604, 606, 608, 610, 612 and 614 have been assigned respectively font attributes based on the association table in FIG. 5 and displayed respectively in the associated font. During the parsing, this module allows sequence selections of data elements based on the reading order of the input document 602 to edit their font information. This module also allows region grouping of data elements to edit their font information. This module can also provide an auxiliary view of the association table.